

(19)



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

EP 0 855 444 A2

(12)

## EUROPEAN PATENT APPLICATION

(43) Date of publication:  
29.07.1998 Bulletin 1998/31

(51) Int Cl.<sup>6</sup>: C12N 15/57, C12N 9/64,  
C12N 1/21, C12N 5/10,  
C07K 16/40, A61K 48/00,  
A61K 39/395, A61K 31/70,  
A61K 38/48, C12Q 1/68,  
C12Q 1/37

(21) Application number: 98300573.7

(22) Date of filing: 27.01.1998

(84) Designated Contracting States:  
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE

(30) Priority: 28.01.1997 GB 9701684

(71) Applicants:

- SMITHKLINE BEECHAM PLC  
Brentford, Middlesex TW8 9EP (GB)
- SMITHKLINE BEECHAM CORPORATION  
Philadelphia Pennsylvania 19103 (US)

(72) Inventors:

- Powell, David J.,  
c/o SmithKline Beecham Pharm.  
King of Prussia, PA 19460 (US)

- Smith, Trudi S., c/o SmithKline Beecham Pharm.  
King of Prussia, PA 19460 (US)
- Chapman, Conrad G.,  
c/o SmithKline Beecham Pharm.  
Harlow, Essex CM19 5AW (GB)
- Murphy, Kay SmithKline Beecham Pharm.  
Harlow Essex CM19 5AW (GB)

(74) Representative: Crump, Julian Richard John et al  
fJ Cleveland,  
40-43 Chancery Lane  
London WC2A 1JQ (GB)

## (54) Aspartic proteinase 2 (ASP2)

(57) ASP2 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing ASP2 polypeptides and polynucle-

otides in the design of protocols for the treatment of Alzheimer's Disease, cancer, and prohormone processing, among others, and diagnostic assays for such conditions.

09 | 47 | 663

- 5
- (ix) TELECOMMUNICATION INFORMATION:
- (A) TELEPHONE: +44 171 405 5875
  - (B) TELEFAX: +44 171 831 0749
  - (C) TELEX:

10 (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2541 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

25	ATGGCCCAAG CCCTGCCCTG GCTCCTGCTG TGAGATGGCG CGGGAGTGCT GCCTGCCAC GGCACCCAGC ACGGCATCCG GCTGCCCTG CGCAGCGGCC TGGGGGGCGC CCCCCCTGGGG CTGCGGCTGC CCCGGGAGAC CGACGAAGAG CCCGAGGAGC CGGGCCGGAG GGGCAGCTTT	60 120 180
30	GTGGAGATGG TGGACAACCT GAGGGGCAAG TCGGGGCAGG GCTACTACGT GGAGATGACC GTGGGCAGCC CCCCGCAGAC GCTAACACATC CTGGTGGATA CAGGCAGCAG TAACTTTGCA GTGGGTGCTG CCCCCCACCC CTTCCCTGCAT CGCTACTACC AGAGGCAGCT GTCCAGCACA TACCGGGACC TCCGGAAAGGG TGTGTATGAG CCCTACACCC AGGGCAAGTG GGAAGGGGAG	240 300 360 420
35	CTGGGCACCG ACCTGGTAAG CATCCCCAT GGCCCCAACG TCACTGTGCG TGCCAAACATT GCTGCATCA CTGAATCAGA CAAGTTCTTC ATCAAACGGCT CCAAACGGGA AGGCATCCTG GGGCTGGCCT ATGCTGAGAT TGCCAGGCCT GACGACTCCC TGGAGCCTTT CTTTGACTCT	480 540 600
40	CTGGTAAAGC AGACCCACGT TCCCAACCTC TTCTCCCTGC AGCTTTGTGG TGCTGGCTTC CCCCCTCAACC AGTCTGAAGT GCTGGCCTCT GTCGGGAGGGA GCATGATCAT TGGAGGTATC GACCACTCGC TGTACACAGG CAGTCTCTGG TATAACACCA TCCGGCGGGA GTGGTATTAT	660 720 780
45	GAGGTGATCA TTGTGCGGGT GGAGATCAAT GGACAGGATC TGAAAATGGA CTGCAAGGAG TACAACATATG ACAAGAGCAT TGTGGACAGT GGCAACACCA ACCTTCGTTT GCCCAAGAAA GTGTTTGAAG CTGCAGTCAA ATCCATCAAG GCAGCCTCCT CCACGGAGAA GTTCCCTGAT	840 900 960
50	GGTTTCTGGC TAGGAGAGCA GCTGGTGTGC TGGCAAGCAG GCACCAACCC TTGGAACATT TTCCCACTCA TCTCACTCTA CCTAATGGGT GAGGTTACCA ACCAGTCCTT CCGCATCACC ATCCTCCGC AGCAATACCT GCGGCCAGTG GAAGATGTGG CCACGTCCA AGACGACTGT	1020 1080 1140
55	TACAAGTTG CCATCTCACA GTCATCCACG GGCACGTGTTA TGGGAGCTGT TATCATGGAG GGCTTCTACG TTGTCTTGA TCGGGCCCGA AAACGAATTG GCTTTGCTGT CAGCGCTTGC CATGTGCACG ATGAGTTCAAG GACGGCAGCG GTGGAAGGCC CTTTTGTCAC CTTGGACATG GAAGACTGTG GCTACAACAT TCCACAGACA GATGAGTCAA CCCTCATGAC CATAGCCTAT GTCATGGCTG CCATCTGCGC CCTCTTCATG CTGCCACTCT GCCTCATGGT GTGTCAGTGG	1200 1260 1320 1380 1440

GAG  
is Glu  
codon

	CGCTGCCTCC	GCTGCCTGCG	CCAGCAGCAT	GATGACTTTG	CTGATGACAT	CTCCCTGCTG	1500
5	AAGTGAGGAG	GCCCCATGGGA	GAAAGATAGA	GATTCCCCTG	GGACCACACC	TCCGTGGTTC	1560
	ACTTTGGTCA	CAAGTAGGAG	ACACAGATGG	CACCTGTGGC	CAGAGCACCT	CAGGACCCCTC	1620
	CCCACCCACC	AAATGCCTCT	GCCTTGATGG	AGAAGGAAAA	GGCTGGCAAG	GTGGGTTCCA	1680
	GGGACTGTAC	CTGTAGGAAA	CAGAAAAGAG	AAGAAAAGAAG	CACTCTGCTG	GCGGGAATAC	1740
10	TCTTGGTCAC	CTCAAATTAA	AGTCGGGAAA	TTCTGCTGCT	TGAAACTTCA	GCCCTGAACC	1800
	TTTGTCCACC	ATTCCCTTAA	ATTCTCCAAC	CCAAAGTATT	CTTCTTTCT	TAGTTTCAGA	1860
	AAGTACTGGCA	TCACACGCAG	GTTACCTTGG	CGTGTGTCCC	TGTGGTACCC	GGGCAGAGAA	1920
	GAGACCAAGC	TTGTTTCCCT	GCTGCCAAA	GTCAGTAGGA	GAGGATGCAC	AGTTTGCTAT	1980
15	TTGCTTTAGA	GACAGGGACT	GTATAAACAA	GCCTAACATT	GGTGCAAAGA	TTGCCTCTTG	2040
	AATTAAAAAA	AAAAACTAGA	TTGACTATTT	ATACAAATGG	GGGCGGCTGG	AAAGAGGAGA	2100
	AGGAGAGGGA	GTACAAAGAC	AGGGAATAGT	GGGATCAAAG	CTAGGAAAGG	CAGAAACACA	2160
	ACCACTCACC	AGTCCTAGTT	TTAGACCTCA	TCTCCAAGAT	ACCATCCCCT	CTCAGAAGAT	2220
20	GGGTGTTGTT	TTCAATGTT	TCTTTCTGT	GGTTGCAGCC	TGACCAAAAG	TGAGATGGGA	2280
	AGGGCTTATC	TAGCCAAAGA	GCTCTTTTTT	AGCTCTCTTA	AATGAAGTGC	CCACTAAGGA	2340
	AGTTCCACTT	GAACACATGG	AATTCTGCC	ATATTAATT	CCATTGTCTC	TATCTGGAAC	2400
	CACCCCTTAA	TCTCTACATA	TGATTAGGTC	CAGCACTTGA	AAATATTCC	AACCNNAAATT	2460
25	TGNCTGGGG	GCTTGCNGN	CCAGGTGCTA	AAAGGGNTTG	GGTAGGNGNC	CNCTNTATN	2520
	TNATNCCTNA	AAAGGTTANN	G				2541

## (2) INFORMATION FOR SEQ ID NO:2:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 501 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: protein

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

	Met	Ala	Gln	Ala	Leu	Pro	Trp	Leu	Leu	Leu	Trp	Met	Gly	Ala	Gly	Val
45	1			5				10				15				
	Leu	Pro	Ala	His	Gly	Thr	Gln	His	Gly	Ile	Arg	Leu	Pro	Leu	Arg	Ser
								20				25			30	
	Gly	Leu	Gly	Gly	Ala	Pro	Leu	Gly	Leu	Arg	Leu	Pro	Arg	Glu	Thr	Asp
50									35			40		45		
	Glu	Glu	Pro	Glu	Glu	Pro	Gly	Arg	Arg	Gly	Ser	Phe	Val	Glu	Met	Val
								50			55		60			
	Asp	Asn	Leu	Arg	Gly	Lys	Ser	Gly	Gln	Gly	Tyr	Tyr	Val	Glu	Met	Thr
55								65			70		75		80	
	Val	Gly	Ser	Pro	Pro	Gln	Thr	Leu	Asn	Ile	Leu	Val	Asp	Thr	Gly	Ser

	85	90	95
5	Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu His Arg Tyr		
	100	105	110
	Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg Lys Gly Val		
	115	120	125
10	Tyr Glu Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu Gly Thr Asp		
	130	135	140
	Leu Val Ser Ile Pro His Pro Asn Val Thr Val Arg Ala Asn Ile		
	145	150	155
15	Ala Ala Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly Ser Asn Trp		
	165	170	175
	Glu Gly Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg Pro Asp Asp		
	180	185	190
20	Ser Leu Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr His Val Pro		
	195	200	205
	Asn Leu Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro Leu Asn Gln		
	210	215	220
25	Ser Glu Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile Gly Gly Ile		
	225	230	235
	Asp His Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro Ile Arg Arg		
	245	250	255
30	Glu Trp Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile Asn Gly Gln		
	260	265	270
	Asp Leu Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys Ser Ile Val		
	275	280	285
35	Asp Ser Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val Phe Glu Ala		
	290	295	300
	Ala Val Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys Phe Pro Asp		
	305	310	315
40	Gly Phe Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala Gly Thr Thr		
	325	330	335
	Pro Trp Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met Gly Glu Val		
	340	345	350
45	Thr Asn Gln Ser Phe Arg Ile Thr Ile Leu Pro Gln Gln Tyr Leu Arg		
	355	360	365
	Pro Val Glu Asp Val Ala Thr Ser Gln Asp Asp Cys Tyr Lys Phe Ala		
	370	375	380
50	Ile Ser Gln Ser Ser Thr Gly Thr Val Met Gly Ala Val Ile Met Glu		
	385	390	395
	Gly Phe Tyr Val Val Phe Asp Arg Ala Arg Lys Arg Ile Gly Phe Ala		
	405	410	415
55	Val Ser Ala Cys His Val His Asp Glu Phe Arg Thr Ala Ala Val Glu		
	420	425	430

Same as in  
Seq ID No: 2  
of 09/471, 669

Gly Pro Phe Val Thr Leu Asp Met Glu Asp Cys Gly Tyr Asn Ile Pro  
 435 440 445

5 Gln Thr Asp Glu Ser Thr Leu Met Thr Ile Ala Tyr Val Met Ala Ala  
 450 455 460

Ile Cys Ala Leu Phe Met Leu Pro Leu Cys Leu Met Val Cys Gln Trp  
 465 470 475 480

10 Arg Cys Leu Arg Cys Leu Arg Gln Gln His Asp Asp Phe Ala Asp Asp  
 485 490 495

Ile Ser Leu Leu Lys  
 500

15

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2370 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

30	GGCAGCTTG TGGAGATGGT GGACAACTG AGGGCAAGT CGGGGCAGGG CTACTACGTG	60
	GAGATGACCG TGGGCAGCCC CCCGCAGACG CTCAACATCC TGGTGGATAC AGGCAGCAGT	120
35	AACTTTGCAG TGGGTGCTGC CCCCCACCCC TTCTGCATC GCTACTACCA GAGGCAGCTG	180
	TCCAGCACAT ACCGGGACCT CCGGAAGGGT GTGTATGAGC CCTACACCCA GGGCAAGTGG	240
	GAAGGGGAGC TGGGCACCGA CCTGGTAAGC ATCCCCCATG GCCCCAACGT CACTGTGCGT	300
	GCCAACATTG CTGCCATCAC TGAATCAGAC AAGTTCTTCA TCAACGGCTC CAACTGGAA	360
40	GGCATCCTGG GGCTGGCCTA TGCTGAGATT GCCAGGCCTG ACGACTCCCT GGAGCCTTTC	420
	TTTGAECTC TGGTAAAGCA GACCCACGTT CCCAACCTCT TCTCCCTGCA GCTTTGTGGT	480
	GCTGGCTTCC CCCTCAACCA GTCTGAAGTG CTGGCCTCTG TCGGAGGGAG CATGATCATT	540
	GGAGGTATCG ACCACTCGCT GTACACAGGC AGTCTCTGGT ATACACCCAT CGGGCGGGAG	600
45	TGGTATTATG AGGTGATCAT TGTGGGGTG GAGATCAATG GACAGGATCT GAAAATGGAC	660
	TGCAAGGAGT ACAACTATGA CAAGAGCATT GTGGACAGTG GCACCACCAA CCTTCGTTG	720
	CCCAAGAAAG TGTGGAAGC TGCAGTCAAA TCCATCAAGG CAGCCTCTCC ACGGGAGAAG	780
	TTCCCTGATG GTTCTGGCT AGGAGAGCAG CTGGTGTGCT GGCAAGCAGG CACCACCCCT	840
50	TGGAACATTT TCCCAGTCAT CTCACTCTAC CTAATGGGTG AGGTTACCAA CCAGTCCTTC	900
	CGCATCACCA TCCTTCCGCA GCAATACCTG CGGCCAGTGG AAGATGTGGC CACGTCCCAA	960
	GACGACTGTT ACAAGTTGCA CATCTCACAG TCATCCACCG GCACTGTTAT GGGAGCTGTT	1020
	ATCATGGAGG GCTTCTACGT TGTCTTGAT CGGGCCCGAA AACGAATTGG CTTTGCTGTC	1080
55	AGCGCTTGCC ATGTGCACGA TGAGTCAGG ACGGCAGCGG TGGAAGGCC TTTTGTCAAC	1140
	TTGGACATGG AAGACTGTGG CTACAACATT CCACAGACAG ATGAGTCAAC CCTCATGACC	1200

Xaa Leu Lys Arg Leu Xaa

770

5

## (2) INFORMATION FOR SEQ ID NO:5:

## 10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

## (ii) MOLECULE TYPE: cDNA

## 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GATGAGTTCA GGACGGCAG

19

25

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

40

GGTGCCATAT GTGTCTCC

18

## 45 Claims

1. An isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity over its entire length to a nucleotide sequence encoding the ASP2 polypeptide of SEQ ID NO:2; or a nucleotide sequence complementary to said isolated polynucleotide.
2. The polynucleotide of claim 1 wherein said polynucleotide comprises the nucleotide sequence contained in SEQ ID NO:1 encoding the ASP2 polypeptide of SEQ ID NO:2.
3. The polynucleotide of claim 1 wherein said polynucleotide comprises a nucleotide sequence that is at least 80% identical to that of SEQ ID NO: 1 over its entire length.
4. The polynucleotide of claim 3 which is polynucleotide of SEQ ID NO: 1.

5. The polynucleotide of claim 1 which is DNA or RNA.
6. A DNA or RNA molecule comprising an expression system, wherein said expression system is capable of producing a ASP2 polypeptide comprising an amino acid sequence, which has at least 80% identity with the polypeptide of SEQ ID NO:2 when said expression system is present in a compatible host cell.
7. A host cell comprising the expression system of claim 6.
8. A process for producing a ASP2 polypeptide comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture.
9. A process for producing a cell which produces a ASP2 polypeptide thereof comprising transforming or transfecting a host cell with the expression system of claim 6 such that the host cell, under appropriate culture conditions, produces a ASP2 polypeptide.
10. A ASP2 polypeptide comprising an amino acid sequence which is at least 80% identical to the amino acid sequence of SEQ ID NO:2 over its entire length.
11. The polypeptide of claim 10 which comprises the amino acid sequence of SEQ ID NO:2.
12. An antibody immunospecific for the ASP2 polypeptide of claim 10.
13. A method for the treatment of a subject in need of enhanced activity or expression of ASP2 polypeptide of claim 10 comprising:
- (a) administering to the subject a therapeutically effective amount of an agonist to said polypeptide; and/or  
 (b) providing to the subject an isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity to a nucleotide sequence encoding the ASP2 polypeptide of SEQ ID NO:2 over its entire length; or a nucleotide sequence complementary to said nucleotide sequence in a form so as to effect production of said polypeptide activity *in vivo*.
14. A method for the treatment of a subject having need to inhibit activity or expression of ASP2 polypeptide of claim 10 comprising:
- (a) administering to the subject a therapeutically effective amount of an antagonist to said polypeptide; and/or  
 (b) administering to the subject a nucleic acid molecule that inhibits the expression of the nucleotide sequence encoding said polypeptide; and/or  
 (c) administering to the subject a therapeutically effective amount of a polypeptide that competes with said polypeptide for its ligand, substrate, or receptor.
15. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of ASP2 polypeptide of claim 10 in a subject comprising:
- (a) determining the presence or absence of a mutation in the nucleotide sequence encoding said ASP2 polypeptide in the genome of said subject; and/or  
 (b) analyzing for the presence or amount of the ASP2 polypeptide expression in a sample derived from said subject.
16. A method for identifying compounds which inhibit (antagonize) or agonize the ASP2 polypeptide of claim 10 which comprises:
- (a) contacting a candidate compound with cells which express the ASP2 polypeptide (or cell membrane expressing ASP2 polypeptide) or respond to ASP2 polypeptide; and  
 (b) observing the binding, or stimulation or inhibition of a functional response; or comparing the ability of the cells (or cell membrane) which were contacted with the candidate compounds with the same cells which were not contacted for ASP2 polypeptide activity.
17. An agonist identified by the method of claim 16.